Statistica Sinica Preprint No: SS-2025-0493	
Title	To Adaptively Randomize or to Rerandomize: Comparison of Covariate-adaptive Randomization and Rerandomization
Manuscript ID	SS-2025-0493
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202025.0493
Complete List of Authors	Ziji Qin, Feifang Hu and Yang Liu
Corresponding Authors	Yang Liu
E-mails	yangliu2022@ruc.edu.cn

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0493

To Adaptively Randomize or to Rerandomize: A Comparison of Covariate-Adaptive Randomization and Rerandomization

Ziji Qin, Feifang Hu, and Yang Liu

The George Washington University and Renmin University of China

Abstract: In comparative studies, balancing influential baseline covariates is fundamental to the credible assessment of treatment effects. Two prominent design strategies are covariate-adaptive randomization (CAR) and rerandomization (RR). CAR is typically used in clinical trials in which subjects are enrolled sequentially, whereas RR is formulated for settings in which all subjects are randomized simultaneously. Although both approaches aim to reduce covariate imbalance, their mechanisms differ substantially, which has hindered systematic comparisons between them. This article develops a unified framework for rigorously comparing CAR and RR. We show that RR can match the balance properties of CAR only under increasingly stringent acceptance thresholds. We also study the implications of these thresholds for treatment-effect estimation and for the computational burden of RR. Under such stringent thresholds, the computational cost of RR grows polynomially with sample size but exponentially with feature dimension. Consequently, while RR can effectively balance covari-

Corresponding author (Email: yangliu2022@ruc.edu.cn)

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0493

ates in small to moderate studies, its computational cost escalates rapidly as the sample size and feature dimension increase, limiting its scalability in large trials. In contrast, CAR remains computationally efficient while delivering strong balance guarantees under both sequential and simultaneous enrollment settings. These results clarify the theoretical properties of CAR and RR and provide new insights into their relative merits for the design of randomized controlled trials.

Key words and phrases: Covariate-adaptive randomization; Rerandomization; Covariate imbalance; Treatment effect estimation; Computational cost.

1. Introduction

Balancing influential baseline covariates is fundamental to the design of credible comparative studies, as improved covariate balance enhances the comparability of treatment groups and can increase the efficiency of treatment effect estimation (Rosenberger et al., 2008). The choice of a covariate-balancing design is largely dictated by the enrollment mechanism. Broadly speaking, comparative studies can be classified into two settings: subjects are either enrolled sequentially and require treatment immediately after enrollment, or all subjects are enrolled before treatment assignment and randomized simultaneously. The first setting is common in clinical trials and naturally motivates covariate-adaptive randomization (CAR), which adaptively assigns balanced treatments during sequential enrollment (Taves,

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0493

1974; Pocock and Simon, 1975). The second setting is frequently encountered in social science research and motivates the use of rerandomization (RR), under which candidate treatment allocations are repeatedly generated until a prespecified covariate balance criterion is satisfied (Morgan et al., 2012). Although both approaches are designed to improve covariate balance, they differ substantially in their implementation, computational requirements, and practical applicability. Systematic comparisons of these two designs remain limited, and the regimes under which one approach is preferable to the other are not yet fully understood.

CAR is widely adopted in clinical trials to ensure the balance of the covariates (Bruce et al., 2022). By appropriately specifying the imbalance measure, CAR can accommodate discrete covariates (Hu and Hu, 2012), continuous covariates (Yang et al., 2023; Qin et al., 2024), and multiple treatment arms (Hu et al., 2023). Under mild conditions, the resulting imbalance process can be represented as a recurrent Markov chain, whose magnitude is typically $o_p(n^{1/2})$ and may even attain the $O_p(1)$ rate (Ma et al., 2024). Such improved covariate balance enhances the comparability of treatment groups and can lead to substantial gains in the efficiency of treatment-effect estimation (Ma et al., 2015; Liu and Hu, 2023). Although CAR was originally developed for sequential enrollment settings, it can also

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0493

be applied without modification when all subjects are enrolled simultaneously (Liu et al., 2024), thereby providing a flexible design strategy across a broad range of experimental settings.

It is worth noting that, in clinical trials, "rerandomization" also refers to a recruitment strategy in which patients with recurrent conditions are re-enrolled and rerandomized for separate episodes of the same disease (Kahan et al., 2015). By contrast, the RR design formulated by Morgan et al. (2012) is a covariate-balancing strategy in which treatment assignments for a given set of experimental units are repeatedly generated until a prespecified balance criterion is satisfied. With a sufficiently stringent criterion, RR can achieve near-optimal balance (Wang and Li, 2022) and yield substantial gains in the efficiency of treatment-effect estimation (Li et al., 2018; Li and Ding, 2020). Moreover, the Mahalanobis distance can be modified to accommodate a wide range of covariate structures, including discrete covariates (Johansson and Schultzberg, 2022; Wang et al., 2023) and nonlinear associations (Johansson et al., 2021). Because RR requires baseline covariate information for all units to be available before randomization, it may be impractical in many clinical trials, where patients are enrolled sequentially and require immediate treatment. Nevertheless, RR has been widely used in social-science research; see, for example, Brune et al. (2021)

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0493

and Owen (2023). Because "rerandomization" is already well established in the statistics literature, we use this term throughout this paper to denote the covariate-balancing strategy described above.

Despite these developments, formal comparisons between CAR and RR have remained difficult because the two methods are typically analyzed under different theoretical frameworks: CAR is often studied under super-population asymptotics through a Markov-process formulation, whereas RR is commonly analyzed under finite-population theory (Li et al., 2018). This methodological divide has contributed to different perspectives on their relative merits. While some researchers advocate RR for its conceptual simplicity and transparent balance criterion (Johansson and Schultzberg, 2022), others argue that covariate balance should be enforced directly through the randomization procedure itself. As Athey and Imbens (2017) state,

> *"We recommend simply taking care in the original design so that assignments that correspond to unacceptable balance are ruled out from the outset, rather than ruled out ex post, which complicates inference."*

A unified theoretical framework is therefore needed to clarify when the two approaches achieve comparable covariate balance and when their performance differs in a practically meaningful way.

In this article, we develop a unified theoretical framework for comparing

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0493

CAR and RR. We establish sufficient conditions under which RR attains the same asymptotic order of covariate balance as CAR and characterize the corresponding computational cost through the expected number of randomization draws, N. We show that $\mathbb{E}[N]$ grows polynomially with the sample size n, with an exponent that increases with αq, where α characterizes the stringency of the balance requirement and q is the dimension of the balancing feature vector. Consequently, although RR can attain balance comparable to, or even finer than, that achieved by CAR when n and q are relatively small, its computational burden may become prohibitive in large-scale or feature-rich studies. By contrast, CAR remains computationally scalable while providing strong balance guarantees and can be implemented under either sequential or simultaneous enrollment. Numerical studies illustrate the finite-sample implications of these theoretical results.

The rest of this article is organized as follows. Sections 2 and 3 introduce the framework for the comparison and present the main theoretical results. Section 4 examines practical implications for trial design. Numerical studies are reported in Sections 5 and 6. Concluding remarks are given in Section 7. Proofs and additional results are provided in the Supplementary Material.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0493

2. Framework

Consider a two-arm trial with n patients. For $1 \leq i \leq n$, let T_i denote the treatment assignment, with $T_i = 1$ indicating the assignment to treatment and $T_i = 0$ indicating the assignment to control. The treatment and control sample sizes are $n_1 = \sum_{i=1}^{n} T_i$ and $n_0 = n - n_1$, respectively. For $t \in \{0, 1\}$, let $Y_i(t)$ denote the potential outcome of patient i under treatment t. The observed outcome is $Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0)$. The goal of the trial is to obtain comparable treatment arms and to estimate the average treatment effect $\tau = \mathbb{E}[Y_i(1) - Y_i(0)]$. We consider the difference-in-means estimator $\hat{\tau} = \bar{Y}_1 - \bar{Y}_0$, where $\bar{Y}_t = n_t^{-1} \sum_{i:T_i=t} Y_i$ for $t \in \{0, 1\}$.

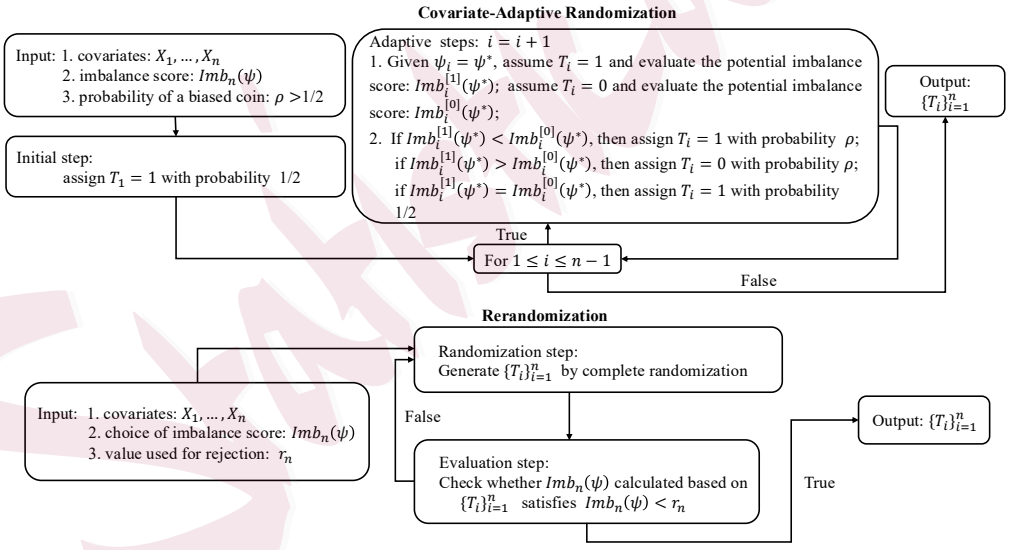

Figure 1: *Flowcharts of CAR and RR.*

Suppose that patient i has a p-dimensional vector of baseline covariates

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0493

$\boldsymbol{X}_i = (X_{i,1}, \ldots, X_{i,p})^\top$, observed prior to assignment. Because the randomization design may involve discretizations or other transformations of $\boldsymbol{X}_i$, let $\psi_i = \psi(\boldsymbol{X}_i)$, where $\psi : \mathbb{R}^p \to \mathbb{R}^q$ represents a feature map that summarizes the covariate information used to assess balance. Define the imbalance vector and the corresponding imbalance score after n assignments as

$$\Lambda_n(\psi) = \sum_{i=1}^{n} (2T_i - 1)\psi_i, \quad \text{and} \quad Imb_n(\psi) = \Lambda_n(\psi)^\top \Lambda_n(\psi). \tag{2.1}$$

The imbalance score in (2.1) provides a common basis for describing CAR and RR. Specifically, one may seek to reduce $Imb_n(\psi)$ as the enrollment proceeds, as in CAR, or by repeatedly generating assignments and retaining one that satisfies a prespecified balance criterion, as in RR. Figure 1 illustrates the two procedures.

Both designs seek to reduce $Imb_n(\psi)$, but they achieve this objective through different mechanisms. CAR is designed for sequential enrollment and improves covariate balance by updating treatment assignment probabilities according to the covariates and treatment assignments of previously enrolled patients. This is often implemented through a biased-coin design (Efron, 1971), which assigns, with probability $\rho > 1/2$, to the treatment that would reduce the current imbalance. In this way, CAR adaptively steers the allocation toward better balance while preserving randomization.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0493

By contrast, RR requires baseline covariate information for the full cohort prior to randomization and is therefore naturally suited to settings in which treatment assignments can be generated simultaneously for a fixed set of patients. It is typically implemented using a Mahalanobis-type imbalance score $Imb_n(\tilde{\psi})$, where $\tilde{\psi}(\boldsymbol{X}) = \Sigma_{\psi\psi}^{-1/2}[\psi(\boldsymbol{X}) - \mu_\psi]$. In practice, when μ_ψ and $\Sigma_{\psi\psi}$ are unknown, they are replaced by their empirical analogues. Given a prespecified threshold r_n, RR repeatedly generates completely randomized assignments until $Imb_n(\tilde{\psi}) < r_n$. The tuning parameters in CAR and RR therefore play distinct roles. ρ in CAR governs the strength of the local preference for assignments that reduce the imbalance, whereas r_n in RR determines the stringency of the global acceptance criterion and is typically chosen as a function of both n and q. More stringent balance requirements, corresponding to smaller values of r_n, admit fewer acceptable allocations and hence entail greater computational burden. In the next section, we compare CAR and RR in terms of their balance properties and discuss their implications for treatment-effect estimation.

3. Main Results

Let "$\sim$" denote asymptotic equivalence: for two positive functions $f(x)$ and $g(x)$, we write $f \sim g$ as $x \to x_0$ if $\lim_{x\to x_0}[g(x)]^{-1}f(x) = 1$. For any p-dimensional vector $\boldsymbol{x}$, let $\|\boldsymbol{x}\| = (\boldsymbol{x}^\top \boldsymbol{x})^{1/2}$ denote the Euclidean norm,

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0493

and let I_p denote the p-dimensional identity matrix. To study the effects of CAR and RR on $\hat{\tau}$, define, for $t \in \{0,1\}$, $m_t(\boldsymbol{X}) = \mathbb{E}[Y(t) \mid \boldsymbol{X}] - \mathbb{E}[Y(t)]$, $\epsilon(t) = Y(t) - \mathbb{E}[Y(t) \mid \boldsymbol{X}]$, and $\Sigma_{Y(t)\psi} = \mathrm{Cov}[Y(t), \psi(\boldsymbol{X})]$. Let $\mu_\psi = \mathbb{E}[\psi(\boldsymbol{X})]$ and $\Sigma_{\psi\psi} = \mathbb{V}[\psi(\boldsymbol{X})]$, and define $\Sigma_{Y\psi} = \Sigma_{Y(1)\psi} + \Sigma_{Y(0)\psi}$, $\sigma_Y^2 = 2[\sigma_{Y(1)}^2 + \sigma_{Y(0)}^2]$. For RR, let N denote the number of randomizations required to obtain an acceptable assignment, and let $F_{\chi_q^2}(x)$ denote the cumulative distribution function of a chi-square random variable with q degrees of freedom. The following assumptions will be used throughout the theoretical developments.

Assumption 3.1. *1.* $\{\boldsymbol{X}_i, Y_i(1), Y_i(0)\}_{i=1}^n$ *are i.i.d. copies of* $\{\boldsymbol{X}, Y(1), Y(0)\}$, *with* $\sigma_{Y(t)}^2 = \mathbb{V}[Y(t)] < \infty$ *and* $\mathbb{E}[\mathbb{V}[Y(t)|\boldsymbol{X}]] > 0$ *for* $t \in \{0,1\}$.

2. The feature map $\psi(\boldsymbol{X})$ *satisfies* $\mathbb{E}[\|\psi(\boldsymbol{X})\|^r] < \infty$ *for some* $r > 5$, $\|\Sigma_{Y(t)\psi}\| < \infty$, *and* $\Sigma_{\psi\psi}$ *is invertible.*

Assumption 3.2. *Suppose that the distribution of* $\psi(\boldsymbol{X})$ *is* Γ_ψ *and that there exists an integer* n_c *and a constant* $1 \geq c_\nu > 0$ *such that* $\Gamma_\psi^{n_c}(A) \geq c_\nu \int_A \nu(y)\, dy$ *for any Borel set* A, *where* $\nu(y)$ *is a density function satisfying* $\inf_{y \in O} \nu(y) > 0$ *for an open set* O, *and* Γ_ψ^k *is the* k*-fold convolution of* Γ_ψ.

Assumption 3.1 provides basic moment conditions on $\{\psi(\boldsymbol{X}), Y(1), Y(0)\}$ needed to establish the theoretical properties of CAR and RR. Assumption 3.2, following Ma et al. (2024), guarantees that the process $\{\Lambda_n(\psi)\}_{n=1}^\infty$

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0493

is ψ-irreducible and underlies the proof that it is Harris positive recurrent (Meyn and Tweedie, 2013).

3.1 Covariate Balance and Treatment Effect Estimation

We first compare the balance properties of CAR and RR in the following proposition (Morgan et al., 2012; Ma et al., 2024).

Proposition 3.1. *1. Under Assumption 3.1 and CAR,* $\mathbb{E}[\|\Lambda_n(\psi)\|^2] = O(n^{1/(c-1)})$ *and thus* $\Lambda_n(\psi) = O_p(n^{1/(2c-2)}) = o_p(n^{1/2})$ *for* $c > 2$. *If Assumption 3.2 also holds,* $\{\Lambda_n(\psi)\}_{n\geq 1}$ *is a positive Harris recurrent Markov chain, and* $\Lambda_n(\psi) = O_p(1)$.

2. Suppose Assumption 3.1 holds and RR is adopted with $r_n = nr$. *Let* $v(r,q) = F_{\chi^2_{q+2}}(r)/F_{\chi^2_q}(r)$ *denote the efficiency factor. Then*

$$n^{-1}Imb_n(\tilde{\psi}) \xrightarrow{\mathbf{D}} \Lambda_{\mathsf{RR}}^\top \Lambda_{\mathsf{RR}} \quad \text{and} \quad n^{-1/2}\Lambda_n(\psi) \xrightarrow{\mathbf{D}} \Sigma_{\psi\psi}^{1/2}\Lambda_{\mathsf{RR}}, \tag{3.1}$$

where $\boldsymbol{Z}_q = (Z_1, \ldots, Z_q)^\top$ *follows* $\mathcal{N}(\mathbf{0}, I_q)$, $\Lambda_{\mathsf{RR}} \overset{\mathbf{D}}{=} \boldsymbol{Z}_q | \{\boldsymbol{Z}_q^\top \boldsymbol{Z}_q < r\}$, *and* Λ_{RR} *satisfies* $\mathbb{E}[\Lambda_{\mathsf{RR}}] = \mathbf{0}_q$ *and* $\mathbb{V}[\Lambda_{\mathsf{RR}}] = v(r,q)I_q$.

Proposition 3.1 compares the balance properties of CAR and RR. Under Assumptions 3.1 and 3.2, CAR yields covariate imbalance $\Lambda_n(\psi)$ of order $O_p(1)$, which corresponds to the optimal rate studied in Hu and Hu (2012) and Ma et al. (2024). Even without Assumption 3.2, CAR still guarantees the slightly weaker rate $\Lambda_n(\psi) = o_p(n^{1/2})$ in general.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0493

For RR, the efficiency factor $v(r,q)$ characterizes the degree of covariate balance achieved under the threshold r. As shown in (3.1), the asymptotic behavior of $Imb_n(\tilde{\psi})$ and $\Lambda_n(\psi)$ is governed by the limiting random vector Λ_{RR}, whose variance is proportional to $v(r,q)$. Smaller values of $v(r,q)$ therefore correspond to finer covariate balance. Theorem 3.3 further establishes that $v(r,q)$ is increasing in the threshold r and decreasing in the feature dimension q. In particular, as $r \downarrow 0$, we have $v(r,q) \downarrow 0$, implying that $n^{-1/2}\Lambda_n(\psi)$ and $n^{-1}Imb_n(\tilde{\psi})$ can be made asymptotically negligible by imposing a sufficiently stringent acceptance criterion. When r is bounded away from zero, however, CAR achieves substantially finer balance.

In summary, when r is sufficiently small, RR can attain balance comparable to, or even finer than, that achieved by CAR. In principle, RR may identify an assignment that minimizes $Imb_n(\psi)$. The key challenge is whether such an assignment can be found with a feasible number of randomization draws N. By contrast, CAR may be viewed as a sequential randomized optimization procedure that provides a computationally feasible way to approximately minimize $Imb_n(\psi)$, thereby maintaining a trade-off between computational cost and the pursuit of optimal covariate balance. We next compare the impact of CAR and RR on $\hat{\tau}$, following the analyses of Li and Ding (2020) and Ma et al. (2024).

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0493

Proposition 3.2. *1. Suppose Assumptions 3.1 and 3.2 hold and CAR is employed. There exists $\sigma_m^2 \geq 0$ such that*

$$n^{-1/2}\Lambda_n(m_1 + m_0) = \tfrac{1}{\sqrt{n}}\textstyle\sum_{i=1}^n (2T_i - 1)[m_1(\boldsymbol{X}_i) + m_0(\boldsymbol{X}_i)] \xrightarrow{\mathbf{D}} \mathcal{N}(0, \sigma_m^2),$$

$$\text{and} \quad \sqrt{n}(\hat{\tau} - \tau) \xrightarrow{\mathbf{D}} \mathcal{N}(0, \sigma_Y^2(1 - \gamma_{Y\psi,\mathsf{CAR}})),$$

where $\gamma_{Y\psi,\mathsf{CAR}} = \sigma_Y^{-2}(\mathbb{V}[m_1(\boldsymbol{X}) + m_0(\boldsymbol{X})] - \sigma_m^2)$.

2. *Suppose Assumption 3.1 holds and RR is implemented with $r_n = nr$. Let Z_0 be a standard normal random variable independent of $\boldsymbol{Z}_q$ defined in Proposition 3.1(ii). Then*

$$\sqrt{n}(\hat{\tau} - \tau) \xrightarrow{\mathbf{D}} \sigma_Y \left(\sqrt{\gamma_{Y\psi,\mathsf{RR}}}\, L_{q,r} + \sqrt{1 - \gamma_{Y\psi,\mathsf{RR}}}\, Z_0\right), \tag{3.2}$$

where $\gamma_{Y\psi,\mathsf{RR}} = \sigma_Y^{-2}\Sigma_{Y\psi}\Sigma_{\psi\psi}^{-1}\Sigma_{\psi Y}$, $L_{q,r} \overset{\mathbf{D}}{=} Z_1 \mid \{\boldsymbol{Z}_q^\top \boldsymbol{Z}_q < r\}$, *and* $L_{q,r}$ *is symmetric about zero with* $\mathbb{V}[L_{q,r}] = v(r, q)$.

Proposition 3.2 shows that CAR and RR affect the distribution of $\hat{\tau}$ in different ways. For CAR, the covariate balance induced by adaptive randomization reduces the variance of $\sqrt{n}(\hat{\tau} - \tau)$, yielding $\sigma_Y^2(1 - \gamma_{Y\psi,\mathsf{CAR}})$. Under complete randomization (CR), the variance of $n^{-1/2}\Lambda_n(m_1 + m_0)$ is $\mathbb{V}[m_1(\boldsymbol{X}) + m_0(\boldsymbol{X})]$, and the variance of $\sqrt{n}(\hat{\tau} - \tau)$ is σ_Y^2. By contrast, balancing $\psi(\boldsymbol{X})$ under CAR generally ensures $0 \leq \sigma_m^2 < \mathbb{V}[m_1(\boldsymbol{X}) + m_0(\boldsymbol{X})]$ (Ma et al., 2024), implying that $\gamma_{Y\psi,\mathsf{CAR}}$ is positive and bounded above by $\sigma_Y^{-2}\mathbb{V}\{m_1(\boldsymbol{X}) + m_0(\boldsymbol{X})\}$. When $m_1(\boldsymbol{X}) + m_0(\boldsymbol{X})$ lies in the linear span of

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0493

$\psi(\boldsymbol{X})$, Proposition 3.1 (i) implies that $\sigma_m^2 = 0$. In this case, CAR achieves the minimal asymptotic variance $2[\mathbb{V}\{\epsilon(1)\}+\mathbb{V}\{\epsilon(0)\}]+\mathbb{V}[m_1(\boldsymbol{X})-m_0(\boldsymbol{X})]$.

By contrast, RR affects $\hat{\tau}$ through the decomposition in (3.2). The first term, $\sigma_Y\sqrt{\gamma_{Y\psi,\mathsf{RR}}}\,L_{q,r}$, represents the component attributable to residual covariate imbalance and depends on the threshold r through $v(r,q)$. As r decreases and $v(r,q) \downarrow 0$, this component vanishes. The second term, $\sigma_Y\sqrt{1-\gamma_{Y\psi,\mathsf{RR}}}\,Z_0$, is orthogonal to $\Lambda_n(\psi)$ and therefore reflects the irreducible variation remaining after balancing $\psi(\boldsymbol{X})$. Thus, in the limiting case of increasingly stringent rerandomization, the minimal asymptotic variance is $\sigma_Y^2(1-\gamma_{Y\psi,\mathsf{RR}})$. Here, $\gamma_{Y\psi,\mathsf{RR}} = \sigma_Y^{-2}\Sigma_{Y\psi}\Sigma_{\psi\psi}^{-1}\Sigma_{\psi Y}$ is closely related to the squared multiple correlation between $\psi(\boldsymbol{X})$ and $Y(1)+Y(0)$ (Morgan et al., 2012), and plays a role analogous to that of $\gamma_{Y\psi,\mathsf{CAR}}$.

These observations highlight two key distinctions between CAR and RR. First, unless r is chosen sufficiently small, the term $\sigma_Y\sqrt{\gamma_{Y\psi,\mathsf{RR}}}\,L_{q,r}$ contributes additional variability under RR, reflecting residual covariate imbalance. Second, even when $v(r,q) \downarrow 0$, the efficiency gains of CAR and RR arise through different mechanisms: the gain under CAR depends on the Markovian behavior of $\Lambda_n(m_1+m_0)$, whereas the gain under RR is determined by the strength of the correlation between $\psi(\boldsymbol{X})$ and $Y(1)+Y(0)$. Despite these differences, both procedures lead to the same qualitative con-

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0493

clusion: the more prognostic information encoded in $\psi(\boldsymbol{X})$, the greater the variance reduction in $\hat{\tau}$.

3.2 Computational Cost of RR

Proposition 3.2 demonstrates that the balance and efficiency gains of RR depend critically on selecting a sufficiently stringent threshold r. It does not, however, quantify the computational burden associated with such a choice. To assess the trade-off between statistical efficiency and computational feasibility, we study the relationship among r, q, $v(r,q)$, and the expected number of randomizations, $\mathbb{E}[N] \approx \{F_{\chi^2_q}(r)\}^{-1}$ by extending Zhou et al. (2018) in the following theorem.

Theorem 3.3. *Let $h(x) = x - 1 - \log x$, and suppose $q \to \infty$. Then the following statements hold.*

1. *Suppose $r = cq + o(1)$ with $c \neq 1$, $c > 0$. (i) If $0 < c < 1$, then $F_{\chi^2_q}(r) \sim [(1-c)\sqrt{\pi q}]^{-1} \exp\{-2^{-1}qh(c)\}$, $\log \mathbb{E}[N] = 2^{-1}qh(c) + 2^{-1}\log(\pi q) + \log(1-c) + o(1)$, and $v(r,q) \to c$. (ii) If $c > 1$, then $F_{\chi^2_q}(r) \sim 1 - [(c-1)\sqrt{\pi q}]^{-1} \exp\{-2^{-1}qh(c)\}$, $\mathbb{E}[N] \to 1$, and $v(r,q) \to 1$.*
2. *Suppose $r = q + o(\sqrt{q})$, then $F_{\chi^2_q}(r) \to \Phi(0)$, $\mathbb{E}[N] \to [\Phi(0)]^{-1} = 2$, and $v(r,q) \to 1$, where $\Phi(x)$ is the cumulative distribution function of a standard normal random variable.*

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0493

3. *Suppose* $r = (q+2)n^{-\alpha} \to 0$ *for some* $\alpha > 0$. *Then* $F_{\chi^2_q}(r) \sim \pi^{-1/2} \exp\{2^{-1}q(1-\alpha\log n) - 2^{-1}\log q + 1\}$, $\log \mathbb{E}[N] = 2^{-1}\alpha q \log n - 2^{-1}q + 2^{-1}\log(\pi q) - 1 + o(1)$, *and* $v(r,q) \sim r/(q+2) = n^{-\alpha} \to 0$.

Theorem 3.3 and Proposition 3.2 together characterize the fundamental trade-off underlying RR. A smaller value of $v(r,q)$, corresponding to finer covariate balance, requires a more stringent rejection threshold r, which reduces the acceptance probability and consequently increases $\mathbb{E}[N]$. For example, if one targets a balance level of the form $v(r,q) \sim n^{-\alpha}$, by choosing $r = (q+2)n^{-\alpha}$, then $\log \mathbb{E}[N] \sim O(\alpha q \log n)$. Thus, the expected computational cost grows polynomially with the sample size n, while the polynomial degree increases linearly with αq. Consequently, increasing α improves covariate balance but also substantially increases the computational burden. In practice, moderate choices such as $\alpha \in (0.4, 0.7)$ often provide a reasonable compromise between computational efficiency and covariate balance. Section 5 illustrates the effect of α on covariate balance, while Section S3.2 of the Supplementary Material examines the corresponding computational cost in more detail. Alternatively, if a fixed level of balance is deemed sufficient, choosing the rejection threshold so that $r/q \to c \in (0,1)$ yields $v(r,q) \to c$, while $\log \mathbb{E}[N] \sim O(q)$. In this case, the computational cost grows exponentially with q, but at a substantially slower rate than under

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0493

the stringent-threshold regime, at the expense of achieving only a fixed level of covariate balance.

It is important to note that computation can quickly become prohibitive when the feature dimension q is large. Enforcing a stringent balance requirement $v(r,q) \sim n^{-\alpha}$ necessitates setting $r \sim (q+2)n^{-\alpha}$, which yields $\mathbb{E}[N]$ of exponential order in $\alpha q \log n$. This burden may be intractable for large n or q. When exact implementation becomes impractical, practitioners may reduce the feature dimension used in the design or relax the target level of $v(r,q)$ by choosing a more moderate threshold. In such settings, CAR may offer a more efficient alternative for achieving covariate balance.

Since $v(r,q) \downarrow 0$ leads to the optimal performance of RR, the following corollary summarizes the corresponding properties of covariate balance and treatment-effect estimation, in line with the findings of Wang and Li (2022).

Corollary 3.1. *Suppose RR uses* $r_n \sim (q+2)n^{1-\alpha}$ *for some* $\alpha > 0$ *and* q *is sufficiently large, then* $v(r,q) \sim n^{-\alpha}$,

$$\Lambda_n(\psi) = o_p(\sqrt{n}), \quad \sqrt{n}(\hat{\tau} - \tau) \xrightarrow{\mathbf{D}} \mathcal{N}\left(0, \sigma_Y^2(1 - \gamma_{Y\psi,\mathsf{RR}})\right),$$

and the number of computations needed satisfies $\log \mathbb{E}[N] \sim O(\alpha \log n \cdot q)$.

By choosing the rejection threshold as $r_n \sim (q+2)n^{1-\alpha}$ for some $\alpha \in (0,1)$, the covariate imbalance $n^{-1/2}\Lambda_n(\psi)$ converges to zero in probability,so that the covariate imbalance is asymptotically negligible at the $\sqrt{n}$

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0493

scale. In this regime, the contribution of $L_{q,r}$ vanishes, and $\hat{\tau}$ is asymptotically normal. While the preceding discussion highlights the distinct mechanisms through which CAR and RR affect treatment-effect estimation, the next corollary formalizes the conditions under which both designs achieve the same minimal asymptotic variance of $\hat{\tau}$ through balancing $\psi(\boldsymbol{X})$.

Corollary 3.2. *Suppose Assumptions 3.1 and 3.2 hold and q is sufficiently large. Under either CAR or RR with $\Lambda_n(\tilde{\psi})$ and $r \sim (q+2)n^{-\alpha}$, we have $n^{-1/2}\Lambda_n(\psi) = o_p(1)$. Furthermore, if $m_1(\boldsymbol{X}) + m_0(\boldsymbol{X}) = [\psi(\boldsymbol{X}) - \mathbb{E}\{\psi(\boldsymbol{X})\}]^\top \boldsymbol{\eta} + m_{\psi_\perp}(\boldsymbol{X})$ for some $\boldsymbol{\eta} \in \mathbb{R}^q$, such that $\mathbb{E}[\boldsymbol{\psi}(\boldsymbol{X}) m_{\psi_\perp}(\boldsymbol{X})] = 0$ and $\mathbb{E}[m_{\psi_\perp}(\boldsymbol{X})] = 0$, then $\gamma_{Y\psi} = \gamma_{Y\psi,\mathsf{CAR}} = \gamma_{Y\psi,\mathsf{RR}} = \sigma_Y^{-2}\boldsymbol{\eta}^\top \Sigma_{\psi\psi}\boldsymbol{\eta}$ and*

$$\sqrt{n}(\hat{\tau} - \tau) \xrightarrow{\mathbf{D}} \mathcal{N}(0, \sigma_Y^2(1 - \gamma_{Y\psi}))$$

holds for both CAR and RR.

Corollary 3.2 delivers two key conclusions. First, RR can attain covariate balance comparable to that of CAR (i.e., $v(r,q) \downarrow 0$) when the threshold is chosen as $r \sim (q+2)n^{-\alpha}$ for $\alpha > 0$. Second, when the decomposition and orthogonality conditions in Corollary 3.2 hold, the effects of covariate balancing under CAR and RR become asymptotically equivalent, implying $\gamma_{Y\psi,\mathsf{CAR}} = \gamma_{Y\psi,\mathsf{RR}}$. In this case, the estimator $\hat{\tau}$ has the same asymptotic distribution under both designs.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0493

4. Applications to Trial Design

We illustrate the use of CAR and RR in clinical trial design using the study of Keller et al. (2000). The trial evaluated the effectiveness of nefazodone (NEF) and the combination of nefazodone with the Cognitive Behavioral–Analysis System of Psychotherapy (NEF+CBT) in patients with chronic nonpsychotic depression. The primary endpoint was the final score on the 24-item Hamilton Rating Scale for Depression (Final-HAMD). Five baseline covariates are available: two discrete variables, i.e., treated current depression (TreatedCD; 1 = Yes, 0 = No) and race (Race; 1 = White, 0 = otherwise), and three continuous variables, i.e., body mass index (BMI), age (Age), and baseline Hamilton Depression Rating Scale (HAMD). For illustration, we set the sample size to 500.

In practice, clinical trial designs often rely on discrete covariates. When continuous covariates are included, they are typically discretized according to prespecified rules. Let the first p_0 components of $\boldsymbol{X}_i$ be discrete and the remaining $p - p_0$ components continuous, denoted by $\boldsymbol{X}_{i,\mathsf{D}}$ and $\boldsymbol{X}_{i,\mathsf{C}}$, respectively. Suppose the j-th discrete covariate $X_{i,j}$ has l_j levels, with values $\{x_j^{s_j} : 1 \le s_j \le l_j\}$. For each continuous covariate $(p_0 + 1) \le j \le p$, let $d_j(\cdot)$ denote the discretization rule such that $d_j(X_{i,j}) = x_j^{s_j}$ when $X_{i,j}$ falls into the s_j-th interval. This yields a total of $\prod_{j=1}^{p} l_j$ strata.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0493

Example 4.1 (CAR for Discrete Covariates)**.** *Hu and Hu (2012) introduced the following imbalance measure for CAR:*

$$Imb_n(\psi) = w_{\mathsf{o}} D_n^2 + \sum_{j=1}^{p} \sum_{s_j=1}^{l_j} w_{\mathsf{m},j} D_n^2(j; s_j) + w_{\mathsf{s}} \sum_{\boldsymbol{s} \in \boldsymbol{S}} D_n^2(\boldsymbol{s}) \tag{4.1}$$

$$\text{and} \quad \psi_i = (\sqrt{w_{\mathsf{o}}}, \underbrace{\ldots, \sqrt{w_{\mathsf{m},j}} \mathbb{I}_i\{j; s_j\}, \ldots}_{\sum_{j=1}^{p} l_j \text{ marginal terms}}, \underbrace{\ldots, \sqrt{w_{\mathsf{s}}} \mathbb{I}_i\{s_1, \ldots, s_p\}, \ldots}_{\prod_{j=1}^{p} l_j \text{ stratum terms}})^\top, \tag{4.2}$$

where $\boldsymbol{S}$ is the set of all possible strata $\boldsymbol{s}$, and D_n, $D_n(j; s_j)$, and $D_n(\boldsymbol{s})$ denote the overall, marginal, and within-stratum imbalances, respectively. The nonnegative weights w_{o}, $w_{\mathsf{m},j}$, and w_{s} are associated with the three levels of imbalance satisfying $w_{\mathsf{o}} + \sum_{j=1}^{p} w_{\mathsf{m},j} + w_{\mathsf{s}} = 1$. The indicators $\mathbb{I}_i\{j; s_j\}$ and $\mathbb{I}_i\{s_1, \ldots, s_p\}$ specify whether $X_{i,j} = x_j^{s_j}$, and whether $\boldsymbol{X}_i$ falls into the stratum $(x_1^{s_1}, \ldots, x_p^{s_p})^\top$.

Several covariate-adaptive randomization procedures arise as special cases of the CAR defined by (4.2). When $w_{\mathsf{s}} = 1$, the procedure reduces to the stratified biased-coin design of Baldi Antognini and Zagoraiou (2011), which targets within-stratum imbalance only. When $w_{\mathsf{o}} = w_{\mathsf{s}} = 0$, it coincides with the minimization procedure (Taves, 1974; Pocock and Simon, 1975), focusing exclusively on marginal imbalance. By Proposition 3.1(i), if $w_{\mathsf{m},j} > 0$, then $D_n(j; s_j) = O_p(1)$, and if $w_{\mathsf{s}} > 0$, then $D_n(\boldsymbol{s}) = O_p(1)$, ensuring an adequate level of covariate balance.

Example 4.2 (RR for Discrete Covariates)**.** *Using RR to balance discrete*

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0493

covariates is slightly more complicated. The feature ψ *in* (4.2) *cannot be directly used, since* $D_n(j; s_j)$ *and* $D_n(\boldsymbol{s})$ *are linearly dependent and* $\Sigma_{\psi\psi}$ *is singular. When within-stratum imbalance is prioritized, one may remove the linear dependence by treating one stratum as the reference level, namely,*

$$\psi_i = (\mathbb{I}\{1, \ldots, 1\}, \ldots, \mathbb{I}\{l_1, \ldots, l_p - 1\})^\top , \tag{4.3}$$

where the last stratum $\{l_1, \ldots, l_p\}$ *is treated as the baseline. We refer to this design as stratified RR. Note that the number of strata* $\prod_{j=1}^{p} l_j$ *grows exponentially in* p*, so balancing at the stratum level alone may be insufficient to ensure adequate overall covariate balance when* p *is relatively large.*

In such settings, marginal balancing is often preferred (Taves, 1974; Pocock and Simon, 1975), leading to the feature specification:

$$\psi_i = (\mathbb{I}\{1; 1\}, \ldots, \mathbb{I}\{1; l_1 - 1\}, \ldots, \mathbb{I}\{p; 1\}, \ldots, \mathbb{I}\{p; l_p - 1\})^\top, \tag{4.4}$$

where, for each covariate, the last level $(j; l_j)$ *is treated as the baseline, and we refer to this design as marginal RR. With an appropriate choice of* $r \sim (q+2)n^{-\alpha}$*, where* $q = \prod_{j=1}^{p} l_j - 1$ *for stratified RR and* $q = \sum_{j=1}^{p} l_j - p$ *for marginal RR, the two procedures can achieve covariate balance comparable to that attained by stratified and marginal CAR, respectively.*

Nevertheless, the feature dimensions $\prod_{j=1}^{p} l_j - 1$ *and* $\sum_{j=1}^{p} l_j - p$ *increase with* p*, resulting in a potentially substantial computational burden. In the*

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0493

study of Keller et al. (2000), discretizing the three continuous covariates into two levels leads to 32 *strata and* 10 *margins. As a consequence, stratified RR can be computationally inefficient, while marginal RR may still require a nontrivial trade-off between the degree of balance, governed by the choice of* α*, and the associated computational cost reflected in* N*.*

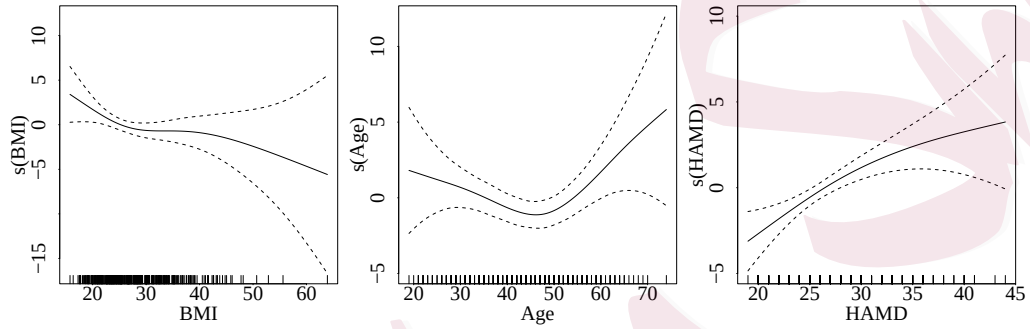

Figure 2: *Estimated nonlinear effect of the covariates under the additive model for the study by Keller et al. (2000).*

Discretizing continuous covariates may induce non-negligible information loss (Ma et al., 2015). For example, the association between X_i and Y_i may be intrinsically nonlinear, as illustrated in Figure 2, and such nonlinear structure is typically obscured by coarse discretization. These considerations motivate the development of designs that explicitly accommodate potential nonlinear covariate–outcome relationships (Ma et al., 2024).

Example 4.3 (CAR and RR for Mixed Discrete and Continuous Covariates)**.** *To balance both discrete and continuous covariates, we adopt* $\psi_i = (\psi_{i,\mathsf{D}}^\top, \psi_{i,\mathsf{C}}^\top)^\top$ *for implementing CAR, where*

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0493

$$\psi_{i,\mathsf{D}} = (\sqrt{w_{\mathsf{o}}},\ \sqrt{w_{\mathsf{m},1}}\mathbb{I}_i\{1;1\}, \ldots, \sqrt{w_{\mathsf{m},p_0}}\mathbb{I}_i\{p_0; s_{p_0}\})^\top, \tag{4.5}$$

$$\psi_{i,\mathsf{C}} = (\sqrt{w_{\mathsf{c},1}}\boldsymbol{X}_{i,\mathsf{C}}^\top, \sqrt{w_{\mathsf{c},2}}\mathrm{vech}(\boldsymbol{X}_{i,\mathsf{C}}\boldsymbol{X}_{i,\mathsf{C}}^\top))^\top. \tag{4.6}$$

Here, $\psi_{i,\mathsf{D}}$ targets marginal imbalance for each component of $\boldsymbol{X}_{i,\mathsf{D}}$, whereas $\psi_{i,\mathsf{C}}$ incorporates both linear and unique quadratic functions of $\boldsymbol{X}_{i,\mathsf{C}}$ to accommodate potential nonlinear covariate–outcome relationships; and $\mathrm{vech}(\cdot)$ *stacks the lower-triangular elements of a symmetric matrix. In practice, $\psi_{i,\mathsf{D}}$ and $\psi_{i,\mathsf{C}}$ may be further augmented with stratum indicators, analogous to* (4.2)*, or with additional basis functions to capture more general forms of nonlinearity. Using this specification of ψ_i ensures that $D_n(j; s_j)$, $\sum_{i=1}^n (2T_i - 1)\boldsymbol{X}_{i,\mathsf{C}}$, and $\sum_{i=1}^n (2T_i - 1)\mathrm{vech}(\boldsymbol{X}_{i,\mathsf{C}}\boldsymbol{X}_{i,\mathsf{C}}^\top)$ are all $O_p(1)$.*

To implement RR, we replace $\psi_{i,\mathsf{D}}$ with the feature vector for balancing marginal distributions as defined in (4.4)*, while retaining the distinct components of $\psi_{i,\mathsf{C}}$ corresponding to the continuous covariates. We further set the rejection threshold as $r \sim (q+2)n^{-\alpha}$ with a moderately large value of α (e.g., $\alpha = 0.5$) to ensure adequate balance. In the study of Keller et al. (2000), this specification yields $q = 8$, resulting in $\alpha q \log n \approx 25$, which can still lead to a nontrivial computational cost.*

In summary, both CAR and RR can be employed to balance discrete

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0493

and continuous covariates. With an appropriate specification of ψ_i, the two procedures can be designed to target the same imbalance measure. In the following sections, we compare CAR and RR in terms of covariate balance, the efficiency of treatment-effect estimation, and computational cost.

5. Numerical Studies

We conduct simulation studies to compare CR, CAR, and RR in terms of covariate balance and the performance of $\hat{\tau} = \bar{Y}_1 - \bar{Y}_0$. Suppose the outcomes are generated from the following linear model:

$$Y_i = T_i\mu_1 + (1 - T_i)\mu_0 + \boldsymbol{\xi}_{i,\mathsf{D}}^\top \boldsymbol{\beta}_\mathsf{D} + \boldsymbol{\xi}_{i,\mathsf{C}}^\top \boldsymbol{\beta}_\mathsf{C} + \epsilon_i, \tag{5.1}$$

where μ_1 and μ_0 denote the effects of the treatment and control; $\boldsymbol{\xi}_{i,\mathsf{D}}$ and $\boldsymbol{\xi}_{i,\mathsf{C}}$ are vectors of discrete and continuous covariates that are directly associated with Y_i, with corresponding coefficients $\boldsymbol{\beta}_\mathsf{D} = \mathbf{1}_{k_0}$ and $\boldsymbol{\beta}_\mathsf{C} = \mathbf{1}_{k-k_0}$; and ϵ_i are independent and identically distributed (i.i.d.) $\mathcal{N}(0, 1)$. Detailed specifications of $(\boldsymbol{\xi}_{i,\mathsf{D}}^\top, \boldsymbol{\xi}_{i,\mathsf{C}}^\top)^\top$ are provided in the following subsections.

We consider two simulation settings. In Study 1, all prognostic covariates are discrete. In Study 2, both discrete and continuous covariates are present, and nonlinear relationships are allowed. In each setting, we compare CR with several CAR and RR designs that differ in the choice of ψ and r. For the CAR design, we set the biased-coin probability to $\rho = 0.9$.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0493

For each RR design, we consider both a *fixed* threshold, where r is chosen as the 5th percentile of a χ^2_{q+2} distribution, and an *adaptive* threshold of the form $r = (q+2)n^{-\alpha}$, with $\alpha \in \{0.1, 0.3, 0.5, 0.7\}$. We set $n \in \{400, 800\}$ and conduct 10,000 simulation replicates. To limit the computational burden, we cap the number of randomization draws at $N = 200{,}000$. If the balance criterion is not satisfied within this limit, we retain the allocation with the smallest observed imbalance. We also report simulated values of $\mathbb{E}[N]$ and the proportions of simulation replicates that reach the cap. In these cases, the target threshold r is unattainable within the computational budget, and the resulting allocations should be viewed as approximations to the intended RR designs.

5.1 Study 1: Discrete Covariates

In this study, we set $\boldsymbol{\xi}_{i,\mathsf{C}} = \mathbf{0}$ and $\boldsymbol{\xi}_{i,\mathsf{D}} = \boldsymbol{X}_{i,\mathsf{D}} = (X_{i,1}, \ldots, X_{i,k_0})^\top$, where $X_{i,1}, \ldots, X_{i,k_0}$ are i.i.d. Bernoulli(1/2). We consider two values of k_0: (i) $k_0 = 2$ and (ii) $k_0 = 5$, corresponding to settings with a relatively small and a relatively large number of strata, respectively. For CAR, we adopt the HH procedure in (4.2) with weights $w_{\mathsf{o}} = 0.1$, $w_{\mathsf{s}} = 0.3$, and $w_{\mathsf{m},j} = 0.6/k_0$ for $1 \le j \le k_0$. For RR, we consider the marginal RR implemented with ψ_i defined in (4.4), and the stratified RR with ψ_i specified in (4.3).

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0493

The marginal RR and stratified RR with fixed and adaptive thresholds are denoted by $\text{RR}_{\mathsf{m,fix}}$, $\text{RR}_{\mathsf{s,fix}}$, $\text{RR}_{\mathsf{m,adp},\alpha}$, and $\text{RR}_{\mathsf{s,adp},\alpha}$, respectively. To facilitate the comparison, we take $\psi_i = (X_{i,1}, \ldots, X_{i,k_0})^\top$ when computing $Imb_n(\psi)$, and report $\sum_{i=1}^n (2T_i - 1)X_{i,1}$ for the marginal imbalance.

Table 1: Evaluation of covariate imbalance and treatment effect estimation under different design schemes for Study 1.

k_0	Design	$Imb_n(\psi)$		$\sum_{i=1}^n (2T_i-1)X_{i,1}$		$\hat{\tau}$	
	n	400	800	400	800	400	800
		mean	mean	mean/s.d.	mean/s.d.	mean/$\sqrt{n}$s.d.	mean/$\sqrt{n}$s.d.
2	CR	399.47	794.29	-0.05/14.19	0.10/19.93	1.00/2.48	1.00/2.44
	HH	2.11	2.11	-0.07/ 1.03	-0.06/ 1.03	1.00/1.99	1.00/2.01
	$\text{RR}_{\mathsf{m,fix}}$	5.49	10.59	0.02/ 1.66	0.02/ 2.30	1.00/2.02	1.00/2.00
	$\text{RR}_{\mathsf{m,adp},0.1}$	90.60	171.61	-0.08/ 6.77	-0.04/ 9.27	1.00/2.25	1.00/2.21
	$\text{RR}_{\mathsf{m,adp},0.3}$	31.75	51.78	0.01/ 4.01	-0.01/ 5.10	1.00/2.11	1.00/2.03
	$\text{RR}_{\mathsf{m,adp},0.5}$	10.17	14.12	0.01/ 2.26	-0.01/ 2.65	1.00/2.04	1.00/2.00
	$\text{RR}_{\mathsf{m,adp},0.7}$	3.11	3.14	0.01/ 1.25	0.01/ 1.26	1.00/2.01	1.00/2.01
	$\text{RR}_{\mathsf{s,fix}}$	13.66	27.70	-0.01/ 2.62	0.03/ 3.74	1.00/2.03	1.00/2.01
	$\text{RR}_{\mathsf{s,adp},0.1}$	93.10	172.50	-0.05/ 6.85	0.03/ 9.29	1.00/2.28	1.00/2.22
	$\text{RR}_{\mathsf{s,adp},0.3}$	31.67	51.85	0.03/ 3.99	0.03/ 5.13	1.00/2.09	1.00/2.06
	$\text{RR}_{\mathsf{s,adp},0.5}$	9.59	13.75	-0.04/ 2.18	-0.02/ 2.62	1.00/2.02	1.00/2.03
	$\text{RR}_{\mathsf{s,adp},0.7}$	2.94	3.28	0.02/ 1.20	0.00/ 1.27	1.00/1.99	1.00/2.03
5	CR	1004.31	2003.95	0.16/14.09	0.14/19.91	1.00/2.98	1.00/2.99
	HH	13.00	12.95	-0.01/ 1.63	-0.02/ 1.62	1.00/2.01	1.00/2.03
	$\text{RR}_{\mathsf{m,fix}}$	79.14	157.58	-0.05/ 4.00	-0.02/ 5.61	1.00/2.18	1.00/2.15
	$\text{RR}_{\mathsf{m,adp},0.1}$	237.78	446.96	0.07/ 6.88	0.07/ 9.46	1.00/2.55	1.00/2.50
	$\text{RR}_{\mathsf{m,adp},0.3}$	79.73	130.49	-0.03/ 3.99	0.05/ 5.09	1.00/2.20	1.00/2.16
	$\text{RR}_{\mathsf{m,adp},0.5}$	24.55	35.14	-0.01/ 2.20	0.04/ 2.64	1.00/2.08	1.00/2.05
	$\text{RR}_{\mathsf{m,adp},0.7}$	7.40	9.19	-0.02/ 1.22	0.02/ 1.35	1.00/2.02	1.00/2.00
	$\text{RR}_{\mathsf{s,fix}}$	275.03	559.02	-0.05/ 7.42	-0.02/10.54	1.00/2.58	1.00/2.61
	$\text{RR}_{\mathsf{s,adp},0.1}$	259.70	488.36	-0.03/ 7.14	0.05/ 9.88	1.00/2.54	1.00/2.57
	$\text{RR}_{\mathsf{s,adp},0.3}$	121.01	237.64	-0.01/ 4.91	-0.06/ 6.97	1.00/2.28	1.00/2.27
	$\text{RR}_{\mathsf{s,adp},0.5}$	121.91	236.72	0.04/ 4.95	-0.04/ 7.00	1.00/2.29	1.00/2.29
	$\text{RR}_{\mathsf{s,adp},0.7}$	120.93	236.86	0.07/ 4.92	0.04/ 6.92	1.00/2.32	1.00/2.28

Table 1 demonstrates the advantages of HH and RR relative to CR. Compared with CR, both HH and RR substantially improve covariate balance. In particular, under HH, both the mean $Imb_n(\psi)$ and the standard deviation of $\sum_{i=1}^n (2T_i - 1)X_{i,1}$ remain stable as n increases from 400 to 800, regardless of whether k_0 equals 2 or 5. This behavior is consistent with

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0493

the $O_p(1)$ rate implied by Proposition 3.1 (i), and further translates into a noticeable variance reduction for $\hat{\tau}$: the standard deviation of $\sqrt{n}\hat{\tau}$ is close to 2, which is the minimal value obtained when $n^{-1/2}\Lambda_n(\psi) = o_p(1)$.

Table 2: $\mathbb{E}[N]$ and the proportions of cap hits (% Hits) for RR designs in Study 1.

k_0	n		RR_m					RR_s				
			fix	adp, 0.1	adp, 0.3	adp, 0.5	adp, 0.7	fix	adp, 0.1	adp, 0.3	adp, 0.5	adp, 0.7
2	400	$\mathbb{E}[N]$	19	2	4	10	31	21	2	6	34	220
		%Hits	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
	800	$\mathbb{E}[N]$	19	2	4	14	62	20	2	8	54	439
		%Hits	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
5	400	$\mathbb{E}[N]$	20	2	20	301	5,765	23	36	198,876	200,000	200,000
		%Hits	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	98.97%	100.00%	100.00%
	800	$\mathbb{E}[N]$	20	3	30	680	18,593	21	58	199,910	200,000	200,000
		%Hits	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	99.93%	100.00%	100.00%

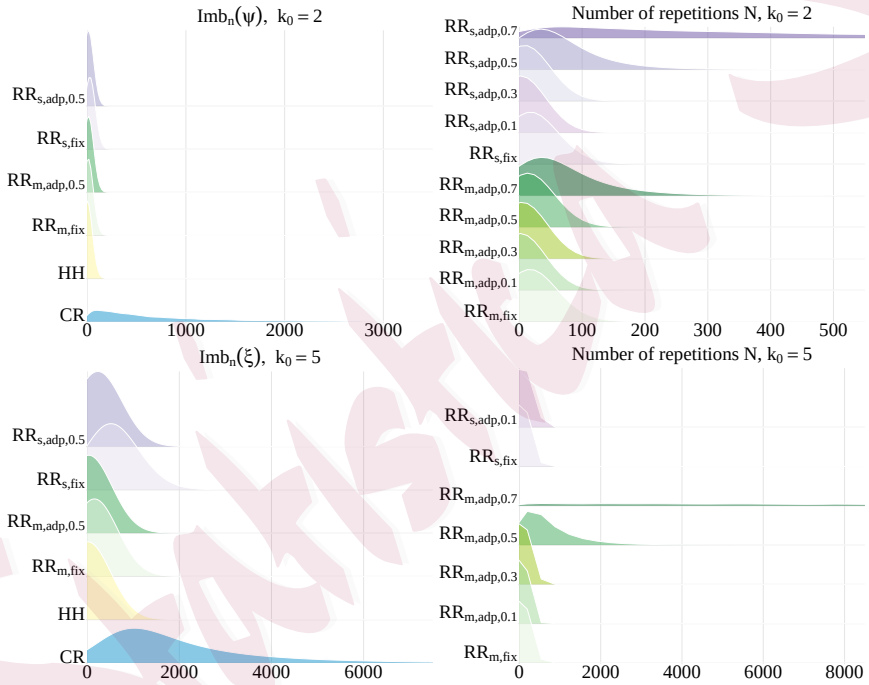

Figure 3: *The distributions of $Imb_n(\psi)$ (left) and N used for RR (right) in Study 1 with $n = 800$. The first and second rows correspond to $k_0 = 2$ and 5, respectively. When $k_0 = 5$, the proportions of cap hits for $RR_{\text{s,adp,0.3}}$, $RR_{\text{s,adp,0.5}}$, and $RR_{\text{s,adp,0.7}}$ are close to 1; their distributions of N are therefore omitted.*

For RR with a fixed threshold, i.e., $\text{RR}_{\text{m,fix}}$ and $\text{RR}_{\text{s,fix}}$, both $Imb_n(\psi)$ and the standard deviation of $\sum_{i=1}^{n}(2T_i - 1)X_{i,1}$ increase with n, indicat-

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0493

ing that the resulting covariate balance is substantially poorer than that achieved by HH. Consequently, the corresponding variance reduction in $\hat{\tau}$ is also more modest than that achieved under HH.

In contrast, RR with $r = (q+2)n^{-\alpha}$ yields substantially finer balance when α is sufficiently large. For $k_0 = 2$, both $\text{RR}_{\mathsf{m,adp},\alpha}$ and $\text{RR}_{\mathsf{s,adp},\alpha}$ exhibit significant improvements in covariate balance as α increases. In particular, $\text{RR}_{\mathsf{s,adp},0.7}$ and $\text{RR}_{\mathsf{m,adp},0.7}$ achieve values of $Imb_n(\psi)$ and standard deviations of $\sum_{i=1}^{n}(2T_i - 1)X_{i,1}$ that are comparable to those under HH, and the standard deviations of $\sqrt{n}\hat{\tau}$ approach 2. Under $k_0 = 5$, the means of $Imb_n(\psi)$ for $\text{RR}_{\mathsf{m,adp},0.7}$ are even smaller than the values under HH.

When $k_0 = 5$, the behaviors of marginal RR and stratified RR diverge markedly. The performance of $\text{RR}_{\mathsf{m,adp},\alpha}$ remains satisfactory and continues to improve as α increases, albeit at the expense of a rapidly growing $\mathbb{E}[N]$. As reported in Table 2, $\mathbb{E}[N]$ for $\text{RR}_{\mathsf{m,adp},0.7}$ reaches 18,593 when $n = 800$. In contrast, RR_{s} performs poorly when the number of strata ($2^5 = 32$) is large. The resulting covariate imbalance is substantially larger than that under $\text{RR}_{\mathsf{m,adp},\alpha}$, while simultaneously requiring a much larger value of N. Even with a moderate threshold (e.g., $\alpha = 0.5$), the targeted level of balance remains difficult to attain: all 10,000 simulation replicates reach the cap of 200,000 randomization draws before satisfying the balance criterion. These

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0493

findings are consistent with Theorem 3.3, which shows that, under stringent balance criteria, the computational cost of RR increases sharply with q. See Section S3.2 of the Supplementary Material for more detail.

5.2 Study 2: Discrete and Continuous Covariates

Suppose $X_{i,1} \sim$ Bernoulli(1/2) and $X_{i,2}, X_{i,3} \sim \mathcal{N}(0,1)$, with $X_{i,1}$, $X_{i,2}$, and $X_{i,3}$ mutually independent. Under the outcome model (5.1), we consider two data-generating mechanisms. In Model 1, $\xi_{i,\mathsf{D}} = X_{i,1}$ and $\xi_{i,\mathsf{C}} = X_{i,2}\exp\{X_{i,2}/2 + X_{i,3}/2\}$. In Model 2, $\xi_{i,\mathsf{D}} = X_{i,1}$ and $\xi_{i,\mathsf{C}} = (X_{i,2},\, X_{i,3}^2,\, X_{i,2}X_{i,3})^\top$. We evaluate designs using $Imb_n(\psi)$ with $\psi_i = (X_{i,1},\, X_{i,2}, X_{i,3}, X_{i,2}^2, X_{i,3}^2, X_{i,2}X_{i,3})^\top$. Consequently, Model 2 is correctly specified with respect to ψ_i, whereas Model 1 is misspecified.

We compare the following two design approaches. *Approach 1* follows the common practice of discretizing continuous covariates. Define $d_2(X_{i,2}) = \mathbb{I}(X_{i,2} > 0)$ and $d_3(X_{i,3}) = \mathbb{I}(X_{i,3} > 0)$. We consider CAR with HH using $(X_{i,1}, d_2(X_{i,2}), d_3(X_{i,3}))$, as well as $\mathrm{RR}_{\mathsf{mar}}$, the marginal RR with an adaptive threshold $\alpha = 0.5$. *Approach 2* directly incorporates continuous features, as discussed in Section 4. Specifically, define $\psi^*_{i,\mathsf{D}}$ as in (4.5) with $p_0 = 1$, and $\psi^*_{i,\mathsf{C}}$ as in (4.6) with $p - p_0 = 2$. We take $(w_{\mathsf{o}}, w_{\mathsf{m}}, w_{\mathsf{c},1}, w_{\mathsf{c},2})^\top = (0.10, 0.20, 0.28, 0.42)^\top$, and denote the result-

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0493

ing design by $\text{CAR}_{\mathsf{mix}}$. For RR, we set $\psi_i = (\psi_{i,\mathsf{D}}, \psi_{i,\mathsf{C}}^\top)^\top$,$\psi_{i,\mathsf{D}} = \mathbb{I}_i(1;1)$, $\psi_{i,\mathsf{C}} = (X_{i,2}, X_{i,3}, X_{i,2}^2, X_{i,3}^2, X_{i,2}X_{i,3})^\top$, and denote the fixed- and adaptive-threshold designs by RR_{fix} and $\text{RR}_{\mathsf{adp},\alpha}$, respectively.

Table 3: Evaluation of covariate imbalance and treatment effect estimation under different design schemes for Study 2.

n	Design	$Imb_n(\psi)$	$\Lambda_n(\psi)$				$\hat{\tau}$	
			$\psi_j = X_1$	$\psi_j = X_2$	$\psi_j = X_2^2$	$\psi_j = X_2X_3$	model 1	model 2
		mean	mean/s.d.	mean/s.d.	mean/s.d.	mean/s.d.	mean/$\sqrt{n}$s.d.	mean/$\sqrt{n}$s.d.
400	CR	3741.89	0.16/14.04	0.01/19.95	0.21/34.57	-0.43/19.85	1.00/5.06	1.00/4.58
	HH	2118.74	-0.07/ 1.25	0.00/12.21	-0.58/28.21	-0.12/15.84	1.00/4.27	1.00/4.04
	RR_{mar}	2320.22	0.00/ 2.19	-0.04/12.61	0.09/28.42	0.36/19.89	1.00/4.37	1.00/4.19
	$\text{CAR}_{\mathsf{mix}}$	58.12	0.03/ 2.34	-0.04/ 2.77	0.05/ 3.63	-0.01/ 3.41	1.00/2.84	1.00/2.16
	RR_{fix}	568.34	-0.02/ 4.36	0.06/ 8.88	-0.25/12.47	0.06/ 8.87	1.00/3.19	1.00/2.73
	$\text{RR}_{\mathsf{adp},0.1}$	1403.70	-0.01/ 6.94	-0.08/13.83	-0.06/19.70	0.14/13.82	1.00/3.95	1.00/3.49
	$\text{RR}_{\mathsf{adp},0.3}$	464.94	-0.03/ 4.01	-0.02/ 7.96	0.00/11.31	0.14/ 8.01	1.00/3.11	1.00/2.61
	$\text{RR}_{\mathsf{adp},0.5}$	143.77	-0.01/ 2.23	0.03/ 4.40	-0.04/ 6.30	0.06/ 4.44	1.00/2.73	1.00/2.23
	$\text{RR}_{\mathsf{adp},0.7}$	43.95	0.00/ 1.22	-0.02/ 2.44	0.05/ 3.50	-0.02/ 2.44	1.00/2.64	1.00/2.08
800	CR	7672.41	0.04/19.88	0.14/28.37	0.15/49.47	-0.11/28.68	1.00/5.05	1.00/4.59
	HH	4279.78	-0.06/ 1.21	-0.04/17.16	-0.90/39.92	0.28/22.15	1.00/4.21	1.00/3.98
	RR_{mar}	4553.83	0.02/ 2.63	-0.42/17.42	0.11/39.91	-0.20/27.58	1.00/4.32	1.00/4.20
	$\text{CAR}_{\mathsf{mix}}$	56.47	0.00/ 2.32	-0.01/ 2.74	-0.09/ 3.47	-0.03/ 3.41	1.00/2.74	1.00/2.07
	RR_{fix}	1126.08	-0.03/ 6.34	0.27/12.47	0.19/17.69	0.10/12.36	1.00/3.19	1.00/2.70
	$\text{RR}_{\mathsf{adp},0.1}$	2641.95	0.08/ 9.44	0.09/19.04	-0.33/27.00	0.33/19.00	1.00/3.82	1.00/3.42
	$\text{RR}_{\mathsf{adp},0.3}$	756.28	0.01/ 5.11	-0.11/10.25	-0.10/14.40	0.18/10.27	1.00/2.97	1.00/2.51
	$\text{RR}_{\mathsf{adp},0.5}$	202.97	0.02/ 2.65	0.05/ 5.32	-0.10/ 7.47	0.00/ 5.29	1.00/2.69	1.00/2.16
	$\text{RR}_{\mathsf{adp},0.7}$	59.51	0.02/ 1.45	0.03/ 2.86	0.02/ 4.05	0.02/ 2.89	1.00/2.58	1.00/2.04

Table 3 reports the performance of the competing designs. When continuous covariates are discretized, both HH and RR_{mar} yield noticeable improvements over CR. However, imbalance in the underlying continuous covariates remains substantial. This is reflected by the relatively large values of $Imb_n(\psi)$, as well as by the considerable variability of $\sum_{i=1}^n (2T_i - 1)\psi_{i,j}$ for continuous features such as $X_{i,2}$, $X_{i,2}^2$, and $X_{i,2}X_{i,3}$. Correspondingly, the variance reduction of $\hat{\tau}$ is modest under both Model 1 and Model 2, indicating that discretization-based designs only partially control the imbalances most relevant for improving estimation precision.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0493

Table 4: $\mathbb{E}[N]$ and the proportions of cap hits (% Hits) for RR designs in Study 2.

Model	n		RR					
			mar	fix	adp, 0.1	adp, 0.3	adp, 0.5	adp, 0.7
Model 1	400	$\mathbb{E}[N]$	34	21	3	35	956	30,896
		%Hits	0.00%	0.00%	0.00%	0.00%	0.00%	0.15%
	800	$\mathbb{E}[N]$	54	21	3	58	2,410	99,383
		%Hits	0.00%	0.00%	0.00%	0.00%	0.00%	20.26%
Model 2	400	$\mathbb{E}[N]$	34	21	3	36	936	30,317
		%Hits	0.00%	0.00%	0.00%	0.00%	0.00%	0.09%
	800	$\mathbb{E}[N]$	54	21	3	59	2,466	100,655
		%Hits	0.00%	0.00%	0.00%	0.00%	0.00%	20.79%

In contrast, the imbalance under $\text{CAR}_{\mathsf{mix}}$ is $O_p(1)$ across both discrete and continuous features, as evidenced by the stability of $Imb_n(\psi)$ and the standard deviations of $\sum_{i=1}^{n}(2T_i - 1)\psi_{i,j}$ across the values of n. This level of balance translates into marked efficiency gains. The standard deviations of $\sqrt{n}\hat{\tau}$ under $\text{CAR}_{\mathsf{mix}}$ are substantially smaller than those under CR, HH, or RR_{mar}, even under the misspecified outcome model in Model 1.

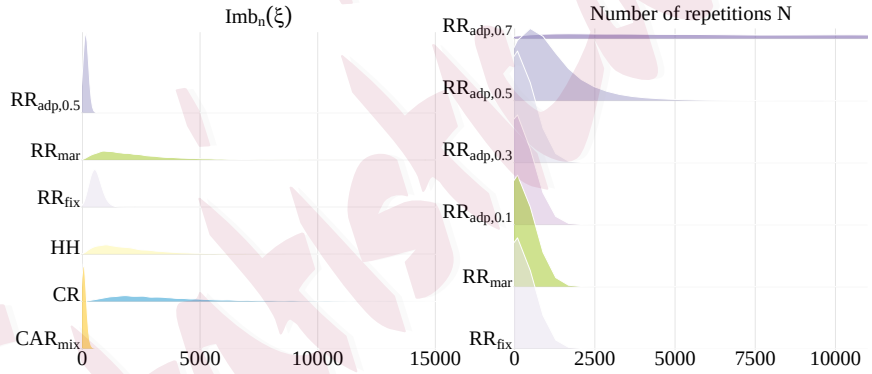

Figure 4: *The distributions of $Imb_n(\psi)$ (left) and N used for RR designs (right) for Study 2 with $n = 800$.*

RR designs based on continuous features exhibit a similar trade-off between covariate balance and computational cost as in Study 1. For RR_{fix}, both the mean of $Imb_n(\psi)$ and the standard deviation of $\sum_{i=1}^{n}(2T_i - 1)\psi_{i,j}$ are smaller than their counterparts under CR, but remain substantially

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0493

larger than those under $\text{CAR}_{\mathsf{mix}}$. When the fixed r is replaced by $r = (q+2)n^{-\alpha}$, $\text{RR}_{\mathsf{adp},\alpha}$ achieves progressively better balance and a smaller variance of $\hat{\tau}$. For $\alpha = 0.5$ and 0.7, the covariate balance attained by $\text{RR}_{\mathsf{adp},\alpha}$ becomes comparable to that achieved by $\text{CAR}_{\mathsf{mix}}$, and the corresponding standard deviations of $\hat{\tau}$ are likewise close to their respective minimum values under both outcome models.

These efficiency gains, however, come at a substantial computational cost. As shown in Table 4, the expected number of randomization draws for $\text{RR}_{\mathsf{adp},0.5}$ is on the order of 10^3 when $n = 400$ and between 10^3 and 10^4 when $n = 800$, with no simulation replicate reaching the cap. By contrast, $\text{RR}_{\mathsf{adp},0.7}$ requires, on average, more than 3×10^4 randomization draws when $n = 400$ and more than 10^5 when $n = 800$, with approximately 20% of the replicates reaching the cap. This pattern is consistent with Theorem 3.3.

In summary, Studies 1 and 2 illustrate the respective strengths and limitations of CAR and RR in balancing covariates. With an appropriate choice of ψ, CAR delivers desirable balance and substantial efficiency gains with negligible computational overhead. In contrast, RR can attain comparable performance only at the cost of potentially prohibitive computation, particularly when n and q are large. Additional supporting results are available in the Supplementary Material.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0493

6. Design Illustration for the Study by Keller et al. (2000)

We revisit the depression trial of Keller et al. (2000) to illustrate the practical implications of our theoretical findings. Because the study includes both discrete and continuous covariates, we compare designs based on covariate discretization with designs that directly incorporate both types of features. We consider implementations of CR, CAR, and RR analogous to those described in Section 5.2. Detailed specifications of the data-generating process and design schemes, together with additional numerical results, are provided in Section S4 of the Supplementary Material. The performance of different designs is evaluated in Tables 5 and 6 and Figure 5.

Table 5 demonstrates the superior performance of CAR for this study. Among the designs based on discretized covariates, HH attains finer balance than RR_{mar} and CR, and yields the smallest standard deviation of $\hat{\tau}$. Nevertheless, the performance of these discretization-based designs remains notably inferior to that of procedures that explicitly leverage continuous features. In particular, $\text{CAR}_{\mathsf{mix}}$ provides a substantial improvement in covariate balance relative to HH. For example, the mean of $Imb_n(\psi)$ decreases from 4257.17 under HH to 136.03 under $\text{CAR}_{\mathsf{mix}}$. This enhanced balancing effect carries over to treatment effect estimation. The empirical standard

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0493

deviation of $\sqrt{n}\hat{\tau}$ is reduced from 5.19 under HH to 4.35 under CAR_{mix}.

Table 5: Evaluation of covariate imbalance and treatment-effect estimation under different design schemes for the study of Keller et al. (2000).

Design	$Imb_n(\psi)$	$\Lambda_n(\psi)$			$\hat{\tau}$
		ψ_j =TreatedCD	ψ_j =BMI	ψ_j = BMI^2	
	mean	mean/s.d.	mean/s.d.	mean/s.d.	mean/$\sqrt{n}$s.d.
CR	7399.96	0.09/15.42	0.38/22.28	0.22/51.06	-4.85/7.03
HH	4257.17	-0.03/ 1.40	-0.14/16.05	-0.63/45.61	-4.85/5.19
RR_{mar}	4370.99	0.00/ 2.35	0.39/16.32	0.72/45.67	-4.85/5.27
CAR_{mix}	136.03	0.00/ 1.92	0.00/ 2.43	-0.02/ 8.19	-4.85/4.35
RR_{fix}	1403.25	0.07/ 5.63	-0.05/11.37	0.03/23.51	-4.85/5.04
$\text{RR}_{\text{adp},0.1}$	2634.52	-0.03/ 7.64	0.17/15.45	0.34/32.59	-4.85/5.65
$\text{RR}_{\text{adp},0.3}$	826.30	-0.04/ 4.34	0.03/ 8.65	0.15/18.20	-4.85/4.72
$\text{RR}_{\text{adp},0.5}$	239.20	0.02/ 2.35	0.04/ 4.70	-0.10/ 9.72	-4.85/4.29
$\text{RR}_{\text{adp},0.7}$	134.07	0.00/ 1.76	-0.01/ 3.50	-0.08/ 7.27	-4.85/4.24

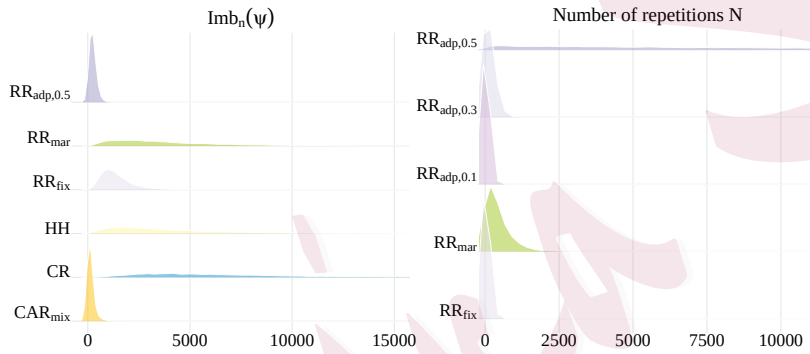

Figure 5: *The distributions of $Imb_n(\psi)$ (left) and N (right) for the study of Keller et al. (2000). The proportion of cap hits for $RR_{\text{adp},0.7}$ is 88.03%; its distribution of N is therefore omitted.*

Table 6: $\mathbb{E}[N]$ and the proportions of cap hits (% Hits) for RR designs in the study by Keller et al. (2000).

	mar	fix	adp,0.1	adp,0.3	adp,0.5	adp,0.7
$\mathbb{E}[N]$	399	21	4	132	12,749	188,205
% Hits	0.00%	0.00%	0.00%	0.00%	0.00%	88.03%

The performance of RR is highly sensitive to the choice of the threshold parameter α. With a fixed or moderately large value (e.g., $\alpha = 0.3$), $\text{RR}_{\text{adp},\alpha}$ yields noticeable improvements over CR by reducing covariate imbalance and achieving modest variance reduction, but its performance remains inferior to that of CAR_{mix}. As α increases (e.g., $\alpha = 0.7$), $\text{RR}_{\text{adp},\alpha}$ can attain a

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0493

level of balance and efficiency comparable to $\text{CAR}_{\mathsf{mix}}$. However, these gains are accompanied by a dramatic increase in computational burden. Tightening the rejection threshold substantially increases the expected number of randomization attempts, particularly when q or n is moderate to large, with $\mathbb{E}[N]$ reaching 188,205 for $\text{RR}_{\mathsf{adp},0.7}$ and the proportion of cap hits reaching 88.03%, thereby limiting the practical feasibility of RR.

7. Conclusion

In this article, we provide a theoretical and empirical comparison of CAR and RR when they target the same imbalance measure. We characterize the conditions under which RR attains covariate balance comparable to that achieved by CAR. Our results show that CAR can achieve the optimal imbalance rate $O_p(1)$, thereby yielding substantial variance reduction in estimating τ. By contrast, the performance of RR is intrinsically tied to the threshold r: a finer balance, and hence greater efficiency gains, can be achieved when r decreases at an appropriate rate. Such tightening, however, entails a rapidly increasing expected computational burden, particularly when the feature dimension q and sample size n are large.

The implications of the theoretical results for the applicability of CAR and RR can be summarized as follows. First, with respect to the enrollment

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0493

scheme, both methods are applicable when all subjects are enrolled before treatment assignment. When subjects are enrolled sequentially, however, RR is generally not applicable, whereas CAR can be implemented as enrollment proceeds. Second, with respect to trial scale, RR is computationally efficient and may even attain optimal balance when both q and n are relatively small. Its computational burden, however, may become substantial as q or n increases. By contrast, CAR is generally more computationally scalable and is therefore better suited to trials involving a larger q or n.

Our work suggests several avenues for future research. First, although our analysis focuses on balance in observed covariates, an important open question is how CAR and RR compare in their ability to mitigate imbalance in unobserved covariates (Liu and Hu, 2022). Second, our investigation underscores the need for more scalable implementations of RR. This motivates the development of sequential or optimization procedures that preserve strong balance while avoiding exponential growth in computational cost. One possible approach is to replace the complete randomization used in RR with CAR, thereby reducing the computational burden and potentially improving covariate balance, as illustrated by Liu et al. (2022). Finally, valid inference following either RR or CAR merits further investigation. Under a correctly specified outcome model, treatment effects may

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0493

be analyzed using regression adjustment that controls for the balancing features ψ (Ma et al., 2020). When the outcome model may be misspecified, however, model-robust procedures are needed to ensure valid inference. A simple approach is to adopt the bootstrap procedure of Shao et al. (2010) to control the Type I error. Alternatively, valid large-sample inference may be constructed through consistent estimation of the asymptotic variance under the corresponding randomization mechanism. Developing such an inferential framework that accommodates both CAR and RR remains an important direction for future research.

Supplementary Material

The Supplementary Material contains proofs of the theoretical results and additional simulation results.

Acknowledgments

The authors would like to thank the editor, the associate editor, and the two reviewers for the insightful comments. Liu's research was supported by the Fundamental Research Funds for the Central Universities and the Research Funds of Renmin University of China (grant no. 202630333).

References

Athey, S. and G. W. Imbens (2017). The econometrics of randomized experiments. In *Handbook of Economic Field Experiments*, Volume 1, pp. 73–140. Amsterdam: Elsevier.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0493

REFERENCES

Baldi Antognini, A. and M. Zagoraiou (2011). The covariate-adaptive biased coin design for balancing clinical trials in the presence of prognostic factors. *Biometrika 98*(3), 519–535.

Bruce, C. L., E. Juszczak, R. Ogollah, C. Partlett, and A. Montgomery (2022). A systematic review of randomisation method use in rcts and association of trial design characteristics with method selection. *BMC Medical Research Methodology 22*, 314.

Brune, L., E. Chyn, and J. Kerwin (2021). Pay me later: Savings constraints and the demand for deferred payments. *American Economic Review 111*(7), 2179–2212.

Efron, B. (1971). Forcing a sequential experiment to be balanced. *Biometrika 58*(3), 403–417.

Hu, F., X. Ye, and L.-X. Zhang (2023). Multi-arm covariate-adaptive randomization. *Science China Mathematics 66*(1), 163–190.

Hu, Y. and F. Hu (2012). Asymptotic properties of covariate-adaptive randomization. *The Annals of Statistics 40*(3), 1794–1815.

Johansson, P., D. B. Rubin, and M. Schultzberg (2021). On optimal rerandomization designs. *Journal of the Royal Statistical Society Series B: Statistical Methodology 83*(2), 395–403.

Johansson, P. and M. Schultzberg (2022). Rerandomization: A complement or substitute for stratification in randomized experiments? *Journal of Statistical Planning and Inference 218*, 43–58.

Kahan, B. C., A. B. Forbes, C. J. Dorés, and T. P. Morris (2015). A re-randomisation design for clinical trials. *BMC Medical Research Methodology 15*, 96.

Keller, M. B., J. P. McCullough, D. N. Klein, et al. (2000). A comparison of nefazodone, the cognitive behavioral-analysis system of psychotherapy, and their combination for the treatment of chronic depression. *New England Journal of Medicine 342*(20), 1462–1470.

Li, X. and P. Ding (2020). Rerandomization and regression adjustment. *Journal of the Royal Statistical Society Series B: Statistical Methodology 82*(1), 241–268.

Li, X., P. Ding, and D. B. Rubin (2018). Asymptotic theory of rerandomization in treatment–

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0493

control experiments. *Proceedings of the National Academy of Sciences 115*(37), 9157–9162.

Liu, Y. and F. Hu (2022). Balancing unobserved covariates with covariate-adaptive randomized experiments. *Journal of the American Statistical Association 117*(538), 875–886.

Liu, Y. and F. Hu (2023). The impacts of unobserved covariates on covariate-adaptive randomized experiments. *The Annals of Statistics 51*(5), 1895–1920.

Liu, Y., Y. Zhou, P. Li, and F. Hu (2022). Adaptive a/b test on networks with cluster structures. In *International Conference on Artificial Intelligence and Statistics*, pp. 10836–10851. PMLR.

Liu, Y., Y. Zhou, P. Li, and F. Hu (2024). Cluster-adaptive network a/b testing: From randomization to estimation. *Journal of Machine Learning Research 25*(170), 1–48.

Ma, W., F. Hu, and L. Zhang (2015). Testing hypotheses of covariate-adaptive randomized clinical trials. *Journal of the American Statistical Association 110*(510), 669–680.

Ma, W., P. Li, L.-X. Zhang, and F. Hu (2024). A new and unified family of covariate adaptive randomization procedures and their properties. *Journal of the American Statistical Association 119*(545), 151–162.

Ma, W., Y. Qin, Y. Li, and F. Hu (2020). Statistical inference for covariate-adaptive randomization procedures. *Journal of the American Statistical Association 115*(531), 1488–1497.

Meyn, S. P. and R. L. Tweedie (2013). *Markov Chains and Stochastic Stability* (3rd ed.). Cambridge, United Kingdom: Cambridge University Press.

Morgan, K. L., D. B. Rubin, et al. (2012). Rerandomization to improve covariate balance in experiments. *The Annals of Statistics 40*(2), 1263–1282.

Owen, S. (2023). College major choice and beliefs about relative performance: An experimental intervention to understand gender gaps in stem. *Economics of Education Review 97*, 102479.

Pocock, S. J. and R. Simon (1975). Sequential treatment assignment with balancing for prog-

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0493

nostic factors in the controlled clinical trial. *Biometrics 31*(1), 103–115.

Qin, Y., Y. Li, W. Ma, H. Yang, and F. Hu (2024). Adaptive randomization via mahalanobis distance. *Statistica Sinica 34*, 353–375.

Rosenberger, W. F., O. Sverdlov, et al. (2008). Handling covariates in the design of clinical trials. *Statistical Science 23*(3), 404–419.

Shao, J., X. Yu, and B. Zhong (2010). A theory for testing hypotheses under covariate-adaptive randomization. *Biometrika 97*(2), 347–360.

Taves, D. R. (1974). Minimization: a new method of assigning patients to treatment and control groups. *Clinical Pharmacology & Therapeutics 15*(5), 443–453.

Wang, X., T. Wang, and H. Liu (2023). Rerandomization in stratified randomized experiments. *Journal of the American Statistical Association 118*(542), 1295–1304.

Wang, Y. and X. Li (2022). Rerandomization with diminishing covariate imbalance and diverging number of covariates. *The Annals of Statistics 50*(6), 3439–3465.

Yang, H., Y. Qin, F. Wang, Y. Li, and F. Hu (2023). Balancing covariates in multi-arm trials via adaptive randomization. *Computational Statistics & Data Analysis 179*, 107642.

Zhou, Q., P. A. Ernst, K. L. Morgan, D. B. Rubin, and A. Zhang (2018). Sequential rerandomization. *Biometrika 105*(3), 745–752.

Ziji Qin
Department of Statistics, The George Washington University
E-mail: zqin123@gwu.edu

Feifang Hu
Department of Statistics, The George Washington University
E-mail: feifang@gwu.edu

Yang Liu
Institute of Statistics and Big Data, Renmin University of China
E-mail: yangliu2022@ruc.edu.cn